

ABSTRACT

Systems and methods for verifying relevance between terms and Web site contents are described. In one aspect, site contents from a bid URL are retrieved. Expanded term(s) semantically and/or contextually related to bid term(s) are calculated. Content similarity and expanded similarity measurements are calculated from respective combinations of the bid term(s), the site contents, and the expanded terms. Category similarity measurements between the expanded terms and the site contents are determined in view of a trained similarity classifier. The trained similarity classifier having been trained from mined web site content associated with directory data. A confidence value providing an objective measure of relevance between the bid term(s) and the site contents is determined from the content, expanded, and category similarity measurements evaluating the multiple similarity scores in view of a trained relevance classifier model.